SENTINETS: User Classification Based on Sentiment for Social Causes within a Twitter Network

Agarwal, Sneha; Guo, Jinlong; Mishra, Shubhanshu; Phelps, Kirstin; Picco, Johna

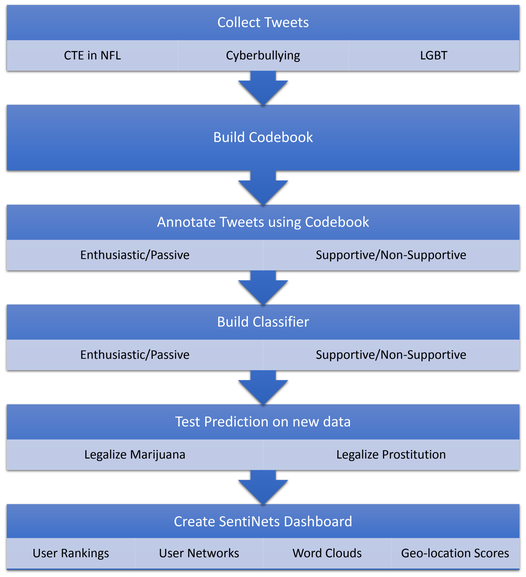
{sagarwa8,jguo24,smishra8,kphelps,picco2}@illinois.edu

Q. how can one leverage large-scale, socio-technical systems toward a smarter society?

A. utilize **sentiment analysis** to improve the identification of **influential** actors within social networks with the creation of a new sentiment classification scheme**.**

Workflow

We set out to solve the defined problem by using the following workflow which will result in finally devising a technique to classify users based on sentiment in a social media network.



Tweet Corpus

1500 Tweets collected using the following social causes as query terms. The corpus didn't have duplicate tweets and had only tweets with length greater than 3 words.

Lesbian Gay Bisexual Transgender [LGBT]

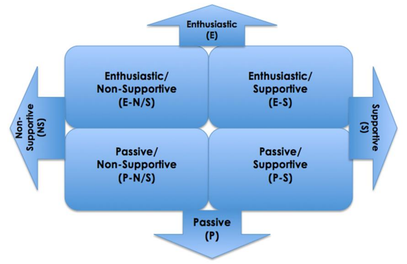
Concussions in National Football League [CTE in NFL]

Cyberbullying

Methods

Classification Schema

Looking at the needs for identifying users/tweets for social causes we found the along side classification of 2 orthogonal classes, to be best suited.
This classification schema allows us to move beyond the positive and negative sentiment classification of tweets to a more audience identification centric approach.



Codebook Generation

We created a codebook which will help us in generating a training corpus for our classifier for each of the following classes. The codebook was used to hand code the 1500 tweet corpus along the 2 orthogonal classification schemas. We avoided using context based knowledge for getting the coding done so as to remove personal opinions from the coding scheme.
For Non-Supportive class we considered the case where the tweets were either directly against the cause or just spreading negative information about the cause. We merged these two cases to build the Non-supportive class as the corpus had very few tweets which were directly against the cause.

Classifier Training

Once we had the training corpus we decided to train a ***Linear Support Vector Machine (SVM)*** based classifier. The classifier was trained using the following features. We used 10 fold cross validation to train the classifier and report the accuracies.

# of Emoticons	***# of URLS***	***# of Mentions***	***# of Hashtags***
Word Features	***# of Double Quotes***	***Length of Tweets***	

Results

Category	Inter Coder Reliability	Accuracy (SVM)
Enthusiastic v/s Passive	93 %	79.0749 %
Supportive v/s Non - Supportive	85 %	76.652 %

coded as negative

"Just watched cyberbully-- it's annoying. Why would she kill herself? It's not worth it. Life is shit so deal with it :P"

Now coded as Enthusiastic & Non-Supportive

coded as positive

"All the best to the retired players suffering from CTE. Spread the word so we can make teh game safer."

Now coded as Enthusiastic & Supportive

coded as positive

"New LGBT Research Study on same sex weddings [link]"

Now coded as Passive & Supportive

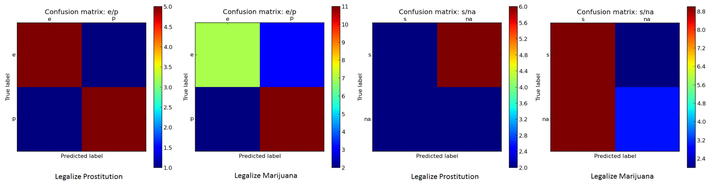
Codebook Results

While building the codebook we observed the following key issues related to classification based on our scales:

- It was found that less than 10% of the people speak openly against a cause in a public platform like social media.
- Supportive/Non-Supportive scale was found to be harder to code consistently as it does require some subjective knowledge as compared to enthusiastic/passive scale

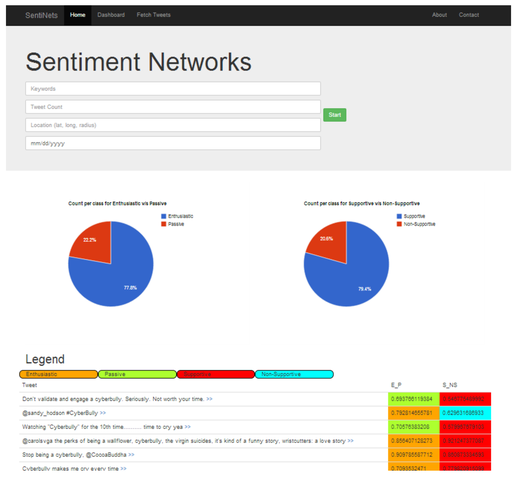
Classifier Testing

We tested our classifier on two new topics (viz. ***"Legalize Marijuana" and "Legalize Prostitution"***) and got very good results for the Enthusiastic v/s Passive Scale. The Supportive v/s Non-Supportive case was more influenced by the nuances in sentiment classification and needs more improvement.



Web Tool

Our web tool allows users to search for tweets for their topic of interest and then show our classification with confidence scores.
The tool also allows users to see the aggregated count of tweets for each class.



Acknowledgement

Faculty Advisor: Jana Diesner

Social Media Expo Team (iConference 2014, Berlin)

Microsoft Fuse Labs

Website

sentinets-smexyyweby.rhcloud.com/coded

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

ILLINOIS